IN THE CLAIMS:

Please CANCEL claims 2 and 8 without prejudice or disclaimer, ADD new claim 15 and AMEND the claims in accordance with the following:

1.   (CURRENTLY AMENDED) An apparatus for extracting information from a formatted document, comprising:

an input unit for inputting a formatted document;

a unit for analyzing the input formatted document and saving analysis results containing particular typographic information;

a unit for identifying special character strings on the basis of the analysis results via preset values of the typographic information, the unit for identifying special character strings determining that a certain character string is special on the basis of the typographic information of the formatted document when the typographic information of the character string is determined to be special typographic information;

a unit for extracting the identified special character strings; and

an output unit for outputting the extracted character strings.

2.   (CANCELLED)

3.   (PREVIOUSLY PRESENTED) The apparatus for extracting information from a formatted document according to claim 1, wherein said formatted document is an HTML document, and said unit for identifying special character strings identifies a certain character string as a special one on the basis of the analysis results with respect to said HTML document when the font size of said character string is determined to be the biggest one among the surrounding character strings.

4.   (PREVIOUSLY PRESENTED) The apparatus for extracting information from a formatted document according to claim 1, wherein said formatted document is an HTML document, and said unit for identifying special character strings determines a certain character string as a special one on the basis of the analysis results with respect to said HTML document when the color and the font of said character string is determined to be a special one among the surrounding character strings.

2

5. (PREVIOUSLY PRESENTED) The apparatus for extracting information from a formatted document according to claim 1, wherein said formatted document is an HTML document, and said unit for identifying special character strings determines a certain character string as a special one on the basis of the analysis results with respect to said HTML document when the font of said character string is determined to be different from the surrounding character strings and the font of said character string is boldface.

6. (PREVIOUSLY PRESENTED) The apparatus for extracting information from a formatted document according to claim 1, wherein said formatted document is an HTML document, and said unit for identifying special character strings determines a certain character string as a special one on the basis of the analysis results with respect to said HTML document when the color of said character string is determined to be different from the surrounding character strings and the font of said character string is boldface.

7. (CURRENTLY AMENDED) A method for extracting information from a formatted document comprising:

inputting a formatted document, analyzing the input formatted document and saving analysis results containing particular typographic information;

identifying special character strings on the basis of the analysis results via preset values of the typographic information;

determining that a certain character string is special on the basis of the typographic information of the formatted document when the typographic information of the character string is determined to be special typographic information;

extracting the identified special character strings; and

outputting the extracted character strings.

8. (CANCELLED)

9. (PREVIOUSLY PRESENTED) The method according to claim 7, wherein said formatted document is an HTML document, and in said identifying of special character strings, a certain character string is determined as a special one on the basis of the analysis results with respect to said HTML document when the font size of said character string is determined to be the biggest one among the surrounding character strings.

3

10. (PREVIOUSLY PRESENTED) The method according to claim 7, wherein said formatted document is an HTML document, and in said identifying of special character strings, a certain character string is determined as a special one on the basis of the analyzing results with respect to said HTML document when the color and the font of said character string is determined to be a special one among the surrounding character strings.

11. (PREVIOUSLY PRESENTED) The method according to claim 7, wherein said formatted document is an HTML document, and in said identifying of special character strings, a certain character string is determined as a special one on the basis of the analysis results with respect to said HTML document when the font of said character string is determined to be different from the surrounding character strings and the font of said character string is boldface.

12. (PREVIOUSLY PRESENTED) The method according to claim 7, wherein said formatted document is an HTML document, and in said identifying of special character strings, a certain character string is determined as a special one on the basis of the analysis results with respect to said HTML document when the color of said character string is determined to be different from the surrounding character strings and the font of said character string is boldface.

13. (PREVIOUSLY PRESENTED) The apparatus for extracting information from a formatted document according to claim 1, wherein the unit for identifying special character strings on the basis of the analysis results sends the typographic information to the unit for extracting the identified special character strings if the typographic information of said character strings is beyond a range of the preset values.

14. (PREVIOUSLY PRESENTED) The method according to claim 7, wherein said extracting extracts the special character strings if the typographic information of said character strings is beyond a range of the preset values.

15. (NEW) The method of claim 7, wherein the special typographic information is the font type and the character string is determined to be special typographic information if the font type differs from the surrounding character strings.